Attorney Docket No. 00280758AA Application Serial No.: 10/749,518

RECEIVED
CENTRAL FAX CENTER
OCT 0 2 2006

Listing of the Claims:

The following is a complete listing of all the claims in the application, with an indication of the status of each:

Claims 1-22 (Canceled).

23. (New) A method of outlier detection comprising:

generating a plurality of synthesized data, each representing a state within a given vector space, said generating including a random number generation;

receiving a plurality of real sample data, each representing a detected real event as represented in said given vector space;

forming a candidate sample set comprising a union of at least a part of said plurality of synthesized data and said plurality of real sample data, said candidate sample set having a starting population, said candidate sample set being unsupervised as to which members will be classified by said method as being outliers;

generating a set of classifiers, each member of said set being a procedure or a representation for a function classifying an operand data as an outlier or a non-outlier, said generating a set of classifiers including:

initializing said set of classifiers to be an empty set,

selectively sampling said candidate sample data to form a learning data set, said selectively sampling including

i) applying said set of classifiers to each of said candidate sample data and, if any classifiers are extant in said set, generating a corresponding set of classification results, and

ii) calculating a sampling probability value for each of said candidate sample data based, at least in part, on the corresponding set of classification results, and

iii) sampling from said candidate data to form said learning data set based, at least in part, on said sampling probability values,

such that said learning data set has a population substantially lower than the starting population,

generating another classifier based on said learning data set,

updating said set of classifiers to include said another classifier, and

repeating said selectively sampling, said generating another classifier, and said updating until said set of classifiers includes at least *t* members; and

generating an outlier detection algorithm based, at least in part, on at least one of said another classifiers, for classifying a datum as being an outlier or a non-outlier.

24. (New) The method of claim 23, further comprising:

receiving a given mis-characterizing cost data associated with at least one of said synthesized samples, representing a cost of said outlier detection algorithm mis-characterizing said at least one synthesized sample as a non-outlier.

25. (New) The method of claim 24, wherein said constructing said learning data set from said candidate sample data is further based on said mis-characterizing cost

26. (New) The method of claim 23, wherein said calculating a sampling probability value includes identifying a consistency within said set of classification results and calculating a probability of said identified consistency using a Binomial probability function.

27. (New) The method of claim 23, wherein said calculating a sampling probability value includes identifying a consistency within said set of classification results and calculating a probability of said identified consistency using a Gaussian probability function.

28. (New) The method of claim 23, said machine-readable code for generating synthesized data generates said synthesized data in accordance with a given statistical likelihood of said generated data meeting an outlier criterion.

29. (New) The method of claim 23, wherein said generating an outlier detection algorithm generates the outlier detection algorithm such that said algorithm applies an aggregate of members of said set of classification algorithms, calculates a corresponding set of detection result data representing each of said aggregate's member's classification, and applies a voting scheme to said corresponding set of detection result data.

30. (New) The method of claim 23, wherein said calculating of said uncertainty value includes identifying a consistency among said set of classification results, and calculating a probability of said consistency, assuming each classification result within said set has a 50-50 probability of representing an operand as meeting said alarm criterion, statistically independent of said operand and of all other classification results within said set.

31. (New) The method of claim 30, wherein said calculating a sampling probability value further includes calculating a probability of said identified consistency using a Binomial probability function.

32. (New) The method of claim 30, wherein said calculating a sampling probability value further includes calculating a probability of said identified consistency using a Gaussian probability function.

33. (New) The method of claim 28, wherein said calculating of said uncertainty value includes identifying a consistency among said set of classification results, and calculating a probability of said consistency, assuming each classification result within said set has a 50-50 probability of representing an operand as

meeting said alarm criteria, statistically independent of said operand and of all other classification results within said set.

34. (New) The method of claim 33, wherein said calculating a sampling probability value further includes calculating a probability of said identified consistency using a Binomial probability function.

35. (New) The method of claim 33, wherein said calculating a sampling probability value further includes calculating a probability of said identified consistency using a Gaussian function.

36. (New) A system for classifying externally detected samples as one of at least normal and an outlier, comprising:

a machine controller having a readable storage medium;

a machine-readable program code, stored on the machine-readable storage medium, having instructions to:

generate a plurality of synthesized data, each representing a state within a given vector space, said generating including a random number generation;

receive a plurality of real sample data, each representing an observed event as represented in said given vector space;

form a candidate sample set comprising a union of at least a part of said plurality of synthesized data and said plurality of real sample data, said candidate sample set having a starting population;

generate a set of classifiers, each member of said set being a procedure or a representation for a function classifying an operand data as a rare event or an outlier, said generating a set of classifiers including:

initializing said set of classifiers to be an empty set,

selectively sampling said candidate sample data to form a learning data set, said selectively sampling including

i) applying said set of classifiers to each of said candidate sample data and, if any classifiers are extant in said set, generating a corresponding set of classification results, and

ii) calculating a sampling probability value for each of said candidate sample data based, at least in part, on the corresponding set of classification results, and

iii) sampling from said candidate data to form said learning data set based, at least in part, on said sampling probability values, such that said learning data set has a population substantially lower than the combined first and second population,

generating another classifier based on said learning data set,

updating said set of classifiers to include said another classifier, and

repeating said selectively sampling, said generating another classifier, and said updating until said set of classifiers includes at least *t* members; and

to generate an outlier detection algorithm based, at least in part, on at least one of said another classifiers, for classifying a datum as being an outlier or a non-outlier.

37. (New) The method of claim 36, wherein said machine-readable program code further comprises instructions for:

receiving a given mis-characterizing cost data associated with at least one of said example abnormal state data, said cost data representing a cost of said outlier detection algorithm mis-characterizing a data equal to said abnormal state data as not meeting said alarm criterion.

38. (New) The system of claim 36, wherein said machine-readable code instructions constructing said learning data set from said candidate sample data include instructions for constructing said learning data based, at least in part, on said mis-characterizing cost.

39. (New) The system of claim 36, wherein said machine-readable code instructions for calculating a sampling probability value include instructions for identifying a consistency within said set of classification results and calculating a probability of said identified consistency using a Binomial probability function.

40. (New) The system of claim 36, wherein said machine-readable code instructions for calculating a sampling probability value include instructions for identifying a consistency within said set of classification results and calculating a probability of said identified consistency using a Gaussian probability function.

41. (New) The system of claim 36, wherein said machine-readable code for generating synthesized data include instruction for generating said data in accordance with a given statistical likelihood meeting an outlier criterion.

42. (New) The system of claim 36, wherein said machine-readable code instructions for generating the outlier detection algorithm include instructions such that said algorithm applies an aggregate of members of said set of classification algorithms, calculates a corresponding set of detection result data representing each of said aggregate's member's classification, and applies a voting scheme to said corresponding set of detection result data.

43. (New) The system of claim 36, wherein said machine-readable code instructions for calculating said uncertainty value include instructions for identifying a consistency among said set of classification results, and calculating a probability of said consistency, assuming each classification result within said set has a 50-50 probability of representing an operand as meeting said alarm criterion, statistically independent of said operand and of all other classification results within said set.

44. (New) The system of claim 36, wherein said machine-readable cod instructions for calculating a sampling probability value include instructions for

calculating a probability of said identified consistency using a Binomial probability function.

45. (New) The system of claim 36, wherein said machine-readable code instructions for calculating a sampling probability value include instructions for calculating a probability of said identified consistency using a Gaussian probability function.

46. (New) The system of claim 41, wherein said machine-readable code instructions for calculating of said uncertainty value include instructions for identifying a consistency among said set of classification results, and calculating a probability of said consistency, assuming each classification result within said set has a 50-50 probability of representing an operand as meeting said alarm criteria, statistically independent of said operand and of all other classification results within said set.

47. (New) The system of claim 45, wherein said machine-readable code instructions for calculating a sampling probability value include instructions for calculating a probability of said identified consistency using a Binomial probability function.

48. (New) The system of claim 45, wherein said machine-readable code instructions for calculating a sampling probability value include instructions for calculating a probability of said identified consistency using a Gausslan probability function.